MATÍAS
SOFTWARE
GROUP

"Sistema de consulta de Indicadores de calidad del aire en ciudades mexicanas"

REPORTE FINAL

Asesoría a cargo de:

Dirección General de Investigación sobre
la Contaminación Urbana y Regional (DGICUR)

Dirección de Investigación sobre la Calidad del Aire (DICA)

Preparado por:

Matías Software Group

Bajo contrato:

INE/I3P-006/2007

Noviembre de 2007

ÍNDICE

Introducción **3**
Justificación **3**
Objetivo **3**
Metodología **4**
La base datos **4**
Los algoritmos utilizados **5**
Teoría de operación **7**
Resultados **8**
Manual del usuario **11**
Conclusiones **11**

Introducción

Resultado del proceso de reingeniería total que se determinó necesario desde las etapas tempranas del proyecto, se rediseñaron y reimplementaron tanto la base de datos como los algoritmos para el cálculo de los indicadores.

En el informe anterior se presentaron las características del nuevo diseño de la estructura general de la base de datos y consideraciones generales sobre los algoritmos a utilizarse.

Es en este Informe Final que se presentan los detalles de la Interfase con el Usuario (UI) desarrollada y los mecanismos con los que se integra con la base de datos. El día 12 de octubre se realizó la primera presentación pública del sistema y se comenzó con el proceso permanente de pruebas.

Durante el desarrollo del sistema, y como se ha hecho énfasis en los informes anteriores, se buscó que se pudiera generalizar tanto como fuera posible la implementación para lograr un sistema extensible. En este Informe Final se detallan los resultados del proceso.

Justificación

Con la reingeniería al sistema de cálculo de indicadores de calidad del aire y la incorporación de la información de las ciudades de Tijuana, Mexicali, Salamanca y otras en el sistema de consulta de indicadores de calidad del aire se evitaran errores humanos durante el procesamiento de datos. Llevar a cabo este proyecto resulta en la generación de indicadores de calidad del aire de las ciudades mexicanas de una manera más rápida y precisa, y al mismo tiempo, permitirá contestar con oportunidad las peticiones de información al respecto. Asimismo, se consolida la automatización del procesamiento de las bases de datos de calidad del aire hasta la generación de indicadores.

Objetivo

Consolidar y adecuar la automatización del procesamiento de las bases de datos de calidad del aire que comenzó en 2006, para generar todos los indicadores que son de interés para la DICA de acuerdo con los criterios establecidos en el INE, e incorporar en el sistema adicional a la información de la Zona Metropiltana del Valle de México (ZMVM), la Zona Metropolitana de Guadalajara (ZMG), la Zona Metropolitana de Monterrey (ZMM), la Zona Metropolitana del Valle de Toluca (ZMVT), Ciudad Juárez y Puebla, la de otras ciudades como son la de Tijuana, Mexicali y Salamanca.

Metodología

La base datos

En el segundo informe parcial se detalló el diseño conceptual de la nueva base de datos.

En este apartado se detallan los objetos que se incorporaron al diseño básico y las razones por las que se crearon.

Conforme se fue desarrollando el sistema se fueron detectando patrones comunes en los “queries” utilizados para el cálculo de los indicadores.

Estos patrones surgen de la clasificación original de los indicadores, primero en tres grupos por el tipo de concentración utilizada: horaria, diaria y de promedio móvil de 8 horas y por otro por el tipo de agregado para las concentraciones diarias: promedio, máximo y máximo de promedio móvil.

Así que para conservar lo ortogonal de la abstracción en la infraestructura se definieron todas esas fuentes de datos independientes de los requerimientos del cálculo de un determinado indicador, por lo que el esquema de la base de datos se fue extendiendo mediante la creación de “vistas” que encapsulan estos patrones comunes.

También se detectó la necesidad de “redensificar” los resultados de los “queries” para el cálculo de algunos indicadores y la forma natural de efectuar ese proceso de redensificación es también mediante el uso de “vistas”.

Por último, para efectos de lograr la máxima velocidad de cálculo posible de algunos indicadores, se requirió la creación de algunas “vista” para optimizar manualmente los “queries”

Se consideró importante, para efectos de la comprensión de los posibles lectores, que la nomenclatura de los objetos en la base de datos tenga un carácter nemónico por lo que se busco también que los nombres de las “vistas” reflejaran su uso.

Por otro lado, la “lógica del negocio” necesitaba reflejarse dentro de la base de datos y eso requería la creación de algunos “tipos de datos” y de algunas funciones y procedimientos dentro de la misma, y para efectos de centralizar esos elementos se desarrolló un “paquete”, llamado “MSG_ACIS”, específico para Oracle 10g.

Las definiciones de todos los objectos que se crearon en la base de datos se agruparon en los siguientes archivos:

Nombre	Contenido
Variables.plsql	Definición de los objetos abstracto “varible” y “contaminante”

Sica.plsql	Definición de “tipos” SQL
Indic.plsql	Definiciones básicas de la base de datos
Msgacispec.plsql	Declaraciónes de paquete “msg_acis”
Msgacisbody.plsql	Cuerpo del paquete “msg_acis”
Views.sql	Vistas empleadas

Como se podrá apreciar de las definiciones, los objetos de la base de datos en si mismos no “calculan” ningún indicador particular, son únicamente “bloques constructivos” de bajo nivel sobre los que se pueden armar “queries” de alto nivel para el cálculo de indicadores específicos.

Este diseño obedece a que se prevé que en el futuro se puedan requerir nuevos indicadores o se modifiquen los procedimientos normados para el calculo de indicadores específicos y se busca que, dado el caso, no sea necesario modificar la estructura de la base de datos.

En el siguiente apartado se describe la capa de abstracción con la que se conecta el sistema de consulta con esta base de datos.

Los algoritmos utilizados

Como se describió en reportes anteriores, el subsistema de consulta y presentación de los indicadores debería ser agnóstico, no sólo respecto al universo de datos disponibles sino también respecto a los indicadores susceptibles de ser calculados. Lo que garantizaría que el incorporar nuevas redes, nuevas estaciones, nuevas variables e incluso nuevos indicadores no requiriera modificación alguna ni en la estructura de la base de datos ni en los programas de consulta.

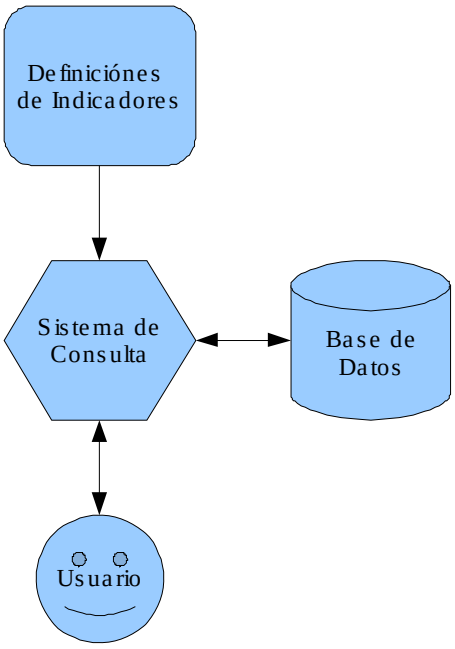
Si para lograr la máxima eficiencia, todo el cálculo para la generación de los indicadores se realiza dentro del motor de la base de datos, el subsistema de consulta y presentación es responsable únicamente de solicitar al usuario los parámetros de su consulta y darle formato y estilo a los reportes generados.

Lograr ese objetivo requirió dos elementos, por una parte todos los metadatos asociados a un determinado contaminante son almacenados en la misma base de datos mediante el objeto “contaminante”.

Pero por otra, se necesitaba implementar un mecanismo que permitiera definir un determinado indicador tanto en términos de su relación con la interfase del usuario, como con el “query” de alto nivel necesario para calcularlo.

Un diagrama del sistema, en grandes bloques, quedaría así:

Figura 1. Diagrama del sistema



Las opciones que el Sistema de Consulta mostrará al Usuario en una interfase genérica estarán determinadas por la definición de los indicadores disponibles y el contenido presente en la base de datos.

El Usuario, con esas opciones seleccionará uno de los parámetros que definen una consulta, con ellos y la definición del indicador seleccionado, el Sistema de Consulta deberá generar dinámicamente el “query” que requiere hacer a la base de datos.

Los resultados del “query” serán presentados al usuario, una vez más, mediante una interfase genérica.

El mecanismo de definición de indicadores necesitaba ser totalmente “declarativo”, estar fuera tanto de la base de datos como del código de la UI, ser fácilmente extensible, ser comprensible para no programadores y estar basado en estándares.

Se definió entonces, utilizando XML, el “Lenguaje de Definición de Indicadores” (LDI), del que se muestra a continuación la DTD (Document Type Definition)[1] completa:

```
<!ELEMENT mode      (indic+)>
<!ATTLIST mode
      id            ID              #REQUIRED
      title         CDATA           #REQUIRED>
<!ELEMENT indic (vvar)>
```

1 Ver: http://es.wikipedia.org/wiki/DTD

```
<!ATTLIST indic
      module        IDREF           #REQUIRED
      desc          CDATA           #REQUIRED
      from          CDATA           #IMPLIED
      norm          (true|false)    "false"
      anual         (true|false)    "false"
      forcest       (true|false)    "false">
<!ELEMENT vvar      (#PCDATA)>
<!ELEMENT modules   (module+)>
<!ELEMENT module    (#PCDATA)>
<!ATTLIST module
      id            ID              #REQUIRED>
```

Mediante LDI se definen, agrupados en “modos”, los “indicadores” disponibles.

La definición de un determinado indicador, no es otra cosa que la asociación entre un conjunto de atributos (normado o no, anual o no, una descripción, las “variables” aplicables, etc.) y un determinado “módulo” de cálculo.

Un “módulo” de cálculo es una “plantilla” para la generación de un “query” en SQL.

En el Anexo VII se presenta completa la Definición de Indicadores actualmente en uso, “calaire.ldi”

Teoría de operación

El Sistema de Consulta está separado en sus partes Cliente y Servidor, utilizando el “Framework GranJefe” que usamos en todos nuestros desarrollos para el WEB.

Al arrancar la aplicación, el Servidor carga las Definiciones de Indicadores y construye, interrogando a la Base de Datos, catálogos dinámicos, ambos son enviados al Cliente.

En el Cliente, mediante AJAX (Asyncronous Javascript And Xml), la interfase HTML es construida dinámicamente conforme el usuario va seleccionando las opciones de su consulta.

Cuando se tienen todos los parámetros requeridos para hacer una consulta, el Cliente los envía al Servidor que, utilizando nuevamente las Definiciones de Indicadores, toma los parámetros recibidos y con el “modulo” correspondiente, construye un “query”, lo envía a la Base de Datos, y sin mayor intermediación devuelve los resultados al Cliente.

El Cliente, entonces, se limita a construir dinámicamente una tabla en que los muestra al Usuario.

De esa manera ni en el Sistema de Consulta (Servidor o Cliente) ni en la Base de Datos se requiere tener “alambrada” información sobre indicadores particulares o su cálculo.

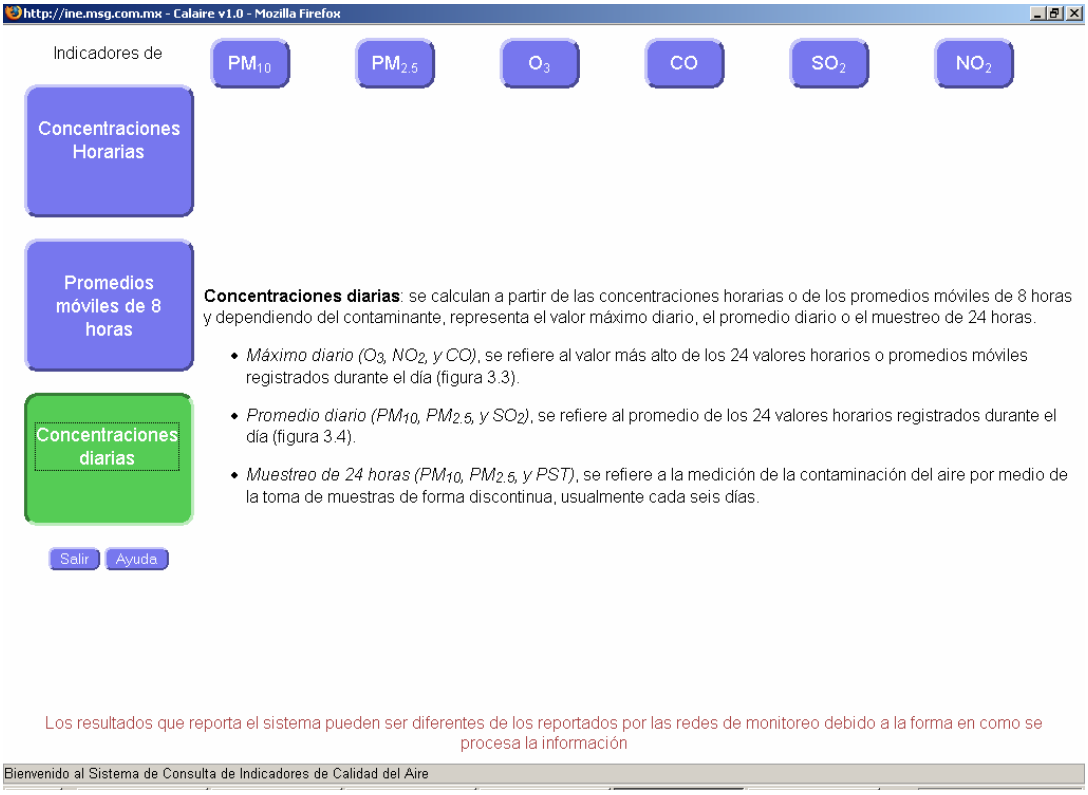
De conformidad con los términos de referencia del proyecto, todo el código del Servidor está desarrollado en el lenguaje de programación PHP, versión 5 y todo el código del Cliente en el lenguaje de programación Javascript.

Resultados

La aplicación de la metodología descrita en la sección anterior da como resultado el *Sistema de consulta de indicadores de calidad del aire*, que estará disponible al público general a través de la página de la Dirección de Investigación sobre Calidad del Aire, después de un tiempo de prueba.

En el *sistema* el usuario puede realizar consultas de indicadores de tres tipos de datos: concentraciones horarias, concentraciones de promedios móviles de ocho horas y concentraciones diarias, de los contaminantes criterio PM_{10}, $PM_{2.5}$, O_3, CO, SO_2 y NO_2. Cuando el usuario posicione el cursor sobre el tipo de concentración o contaminante (botón), se le desplegará al centro de la interfase información del mismo, como se muestra en la siguiente figura.

Interfase del tipo de concentración y contaminantes criterio

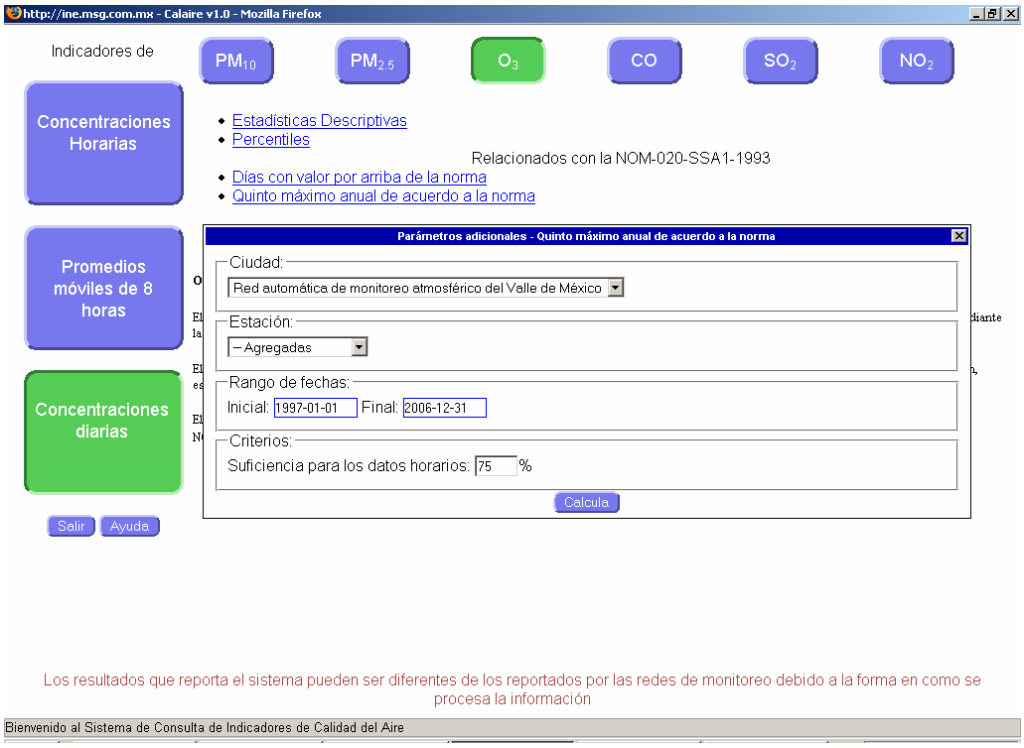

http://ine.msg.com.mx - Calaire v1.0 - Mozilla Firefox
Indicadores de
PM10
PM2.5
O3
CO
SO2
NO2
Concentraciones Horarias
Promedios móviles de 8 horas
Concentraciones diarias
Concentraciones diarias: se calculan a partir de las concentraciones horarias o de los promedios móviles de 8 horas y dependiendo del contaminante, representa el valor máximo diario, el promedio diario o el muestreo de 24 horas.
Máximo diario (O3, NO2, y CO), se refiere al valor más alto de los 24 valores horarios o promedios móviles registrados durante el día (figura 3.3).
Promedio diario (PM10, PM2.5, y SO2), se refiere al promedio de los 24 valores horarios registrados durante el día (figura 3.4).
Muestreo de 24 horas (PM10, PM2.5, y PST), se refiere a la medición de la contaminación del aire por medio de la toma de muestras de forma discontinua, usualmente cada seis días.
Salir
Ayuda
Los resultados que reporta el sistema pueden ser diferentes de los reportados por las redes de monitoreo debido a la forma en como se procesa la información
Bienvenido al Sistema de Consulta de Indicadores de Calidad del Aire

Una vez que el usuario definió el tipo de concentración y contaminante del que quiere hacer su consulta, se desplegará un cuadro de dialogo para que elija los parámetros adicionales, qué ciudad o zona metropolitana, y para la ciudad o zona metropolitana elejida, si se va a llevar a cabo con datos agregados o por estación de monitoreo y el periodo de tiempo (arbitrario, anual, mensual o semanal). Las consultas que se realicen de las concentraciones diarias tienen como parámetro adicional el criterio de suficiencia para calcular el dato base diario con el que se generará el indicador.

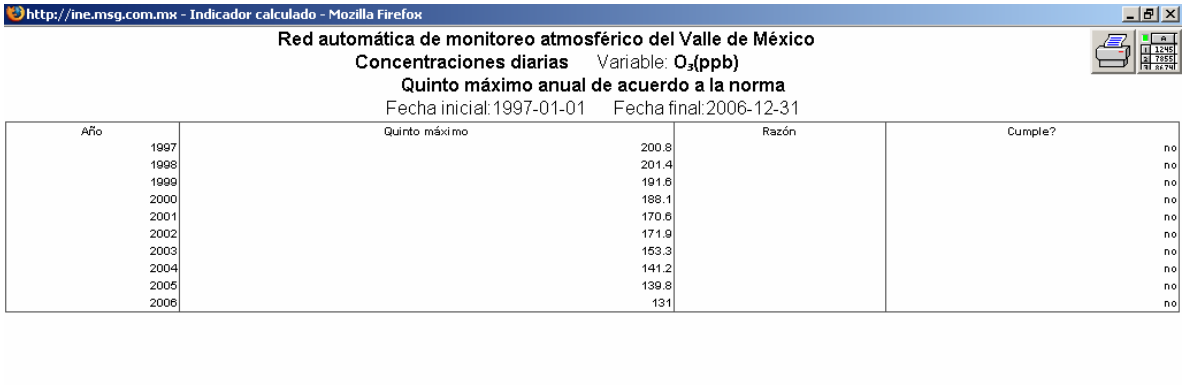
Las consultas se pueden solicitar tanto para indicadores relacionados con las Normas Oficiales Mexicanas (NOM) de calidad del aire (el número de días por arriba del valor de la norma, el percentil 98, los promedios anuales, el quinto máximo, etc.), como para indicadores generales como son las estadísticas descriptivas, percentiles y comportamiento horario.

En la siguientes figuras se muestra un ejemplo de una consulta que se le puede solicitar al sistema y su resultado.

Interfase de una consulta en el Sistema de Indicadores de la calidad del aire en ciudades mexicanas


http://ine.msg.com.mx - Calaire v1.0 - Mozilla Firefox
Indicadores de
PM10
PM2.5
O3
CO
SO2
NO2
Concentraciones Horarias
Estadísticas Descriptivas
Percentiles
Relacionados con la NOM-020-SSA1-1993
Días con valor por arriba de la norma
Quinto máximo anual de acuerdo a la norma
Promedios móviles de 8 horas
Parámetros adicionales - Quinto máximo anual de acuerdo a la norma
Ciudad:
Red automática de monitoreo atmosférico del Valle de México
Estación:
– Agregadas
Rango de fechas:
Inicial: 1997-01-01 Final: 2006-12-31
Criterios:
Suficiencia para los datos horarios: 75 %
Calcula
Concentraciones diarias
Salir
Ayuda
Los resultados que reporta el sistema pueden ser diferentes de los reportados por las redes de monitoreo debido a la forma en como se procesa la información
Bienvenido al Sistema de Consulta de Indicadores de Calidad del Aire

Reporte de una consulta realizada en el Sistema de Indicadores de la calidad del aire en ciudades mexicanas


http://ine.msg.com.mx - Indicador calculado - Mozilla Firefox
Red automática de monitoreo atmosférico del Valle de México
Concentraciones diarias Variable: O3(ppb)
Quinto máximo anual de acuerdo a la norma
Fecha inicial:1997-01-01 Fecha final:2006-12-31

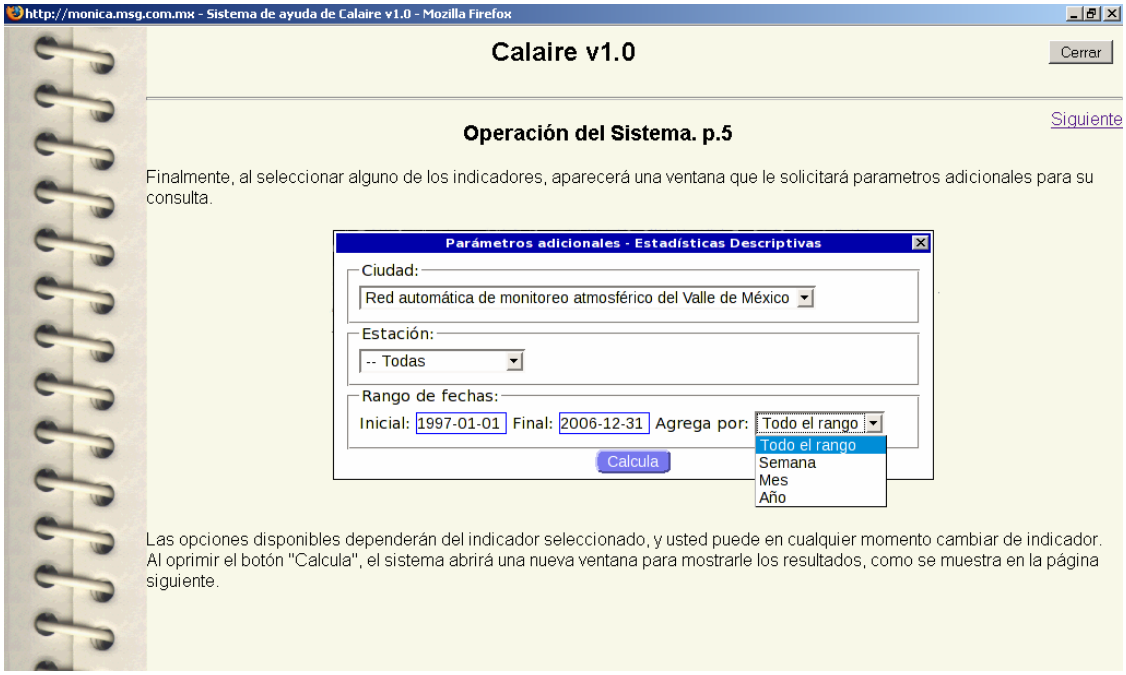
Año	Quinto máximo	Razón	Cumple?
1997	200.8		no
1998	201.4		no
1999	191.6		no
2000	188.1		no
2001	170.6		no
2002	171.9		no
2003	153.3		no
2004	141.2		no
2005	139.8		no
2006	131		no

El indicador que se solicitó en la consulta, para la ZMVM y datos agregados, está relacionado con la NOM-020-SSA1-1993 de ozono (O_3) y corresponde al quinto máximo anual, que se calcula de las concentraciones diarias con un criterio de suficiencia de información de al menos el 75% de los datos horarios para obtener el dato diario. El reporte de la consulta puede mandarse a imprimir o bien guardarse en un archivo de Excel, usando los iconos que se muestran en la parte superior derecha de la pantalla que despliega los resultados.

Manual del usuario

El sistema incorporó el Manual del Usuario en línea para su rápida consulta (y al que se puede acceder a través del icono de ayuda), además la interfase del usuario muestra inmediatamente información relevante a la interacción que el usuario va teniendo con el sistema y en virtud de que toda esa documentación está en archivos simples de HTML, se puede modificar y extender con toda facilidad. En la siguiente figura se muestra un ejemplo de una de las pantallas del manual.

Manual del usuario en linea


http://monica.msg.com.mx - Sistema de ayuda de Calaire v1.0 - Mozilla Firefox
Calaire v1.0
Cerrar
Siguiente
Operación del Sistema. p.5
Finalmente, al seleccionar alguno de los indicadores, aparecerá una ventana que le solicitará parametros adicionales para su consulta.
Parámetros adicionales - Estadísticas Descriptivas
Ciudad:
Red automática de monitoreo atmosférico del Valle de México
Estación:
-- Todas
Rango de fechas:
Inicial: 1997-01-01 Final: 2006-12-31 Agrega por: Todo el rango
Calcula
Todo el rango
Semana
Mes
Año
Las opciones disponibles dependerán del indicador seleccionado, y usted puede en cualquier momento cambiar de indicador. Al oprimir el botón "Calcula", el sistema abrirá una nueva ventana para mostrarle los resultados, como se muestra en la página siguiente.

Conclusiones

Consideramos que los objetivos planteados al arranque del proyecto se han cumplido.

La estructura resultante permite, no solo la extensión del sistema, sino que facilita su mantenimiento y adecuación.

En particular, la separación de las definiciones de los indicadores, permite que éstos se puedan ir ajustando a los requerimientos del área usuaría con total independencia de los elementos de programación del sistema.

En las semanas siguientes se continuarán las pruebas con el sistema hasta su liberación final al público en general. La consulta de indicadores de la calidad del aire podrá realizarse a través de la página de la Dirección de Investigación sobre la Calidad del Aire de la DGICUR.

Con la realización del sistema se logra la automatización del procesamiento de las bases de datos de calidad del aire para generar todos los indicadores que son de interés para la DICA, de acuerdo con los criterios establecidos en el INE para la Zona Metropiltana del Valle de México (ZMVM), la Zona Metropolitana de Guadalajara (ZMG), la Zona Metropolitana de Monterrey (ZMM), la Zona Metropolitana del Valle de Toluca (ZMVT), Ciudad Juárez, Puebla, Tijuana, Mexicali y Salamanca, y evitando de esta manera errores humanos durante el procesamiento de datos, además de generar los indicadores de una manera más rápida y precisa, lo que permitirá contestar con oportunidad las peticiones de información al respecto.

Por nuestra parte, no consideramos al sistema resultante un sistema “acabado” sino una plataforma constructiva para la generación de indicadores presente o futuros, que redundará en una mejor calidad del aire.

No nos resta sino agradecer a la DICA la confianza que se nos ha dado para el desarrollo de este proyecto.

México D.F., noviembre de 2007.